

Directed acyclic graph causal inference framework and its applications in econometrics

Santeri Vääätäjä

School of Business

Bachelor's thesis
Espoo 24.8.2021

Thesis supervisor:

Prof. Pauli Murto

Thesis advisor:

Asst. Prof. Ciprian Domnisoru

Author: Santeri Vääätäjä

Title: Directed acyclic graph causal inference framework and its applications in econometrics

Date: 24.8.2021

Language: English

Number of pages: 4+26

Degree programme: Bachelor's Programme in Economics

Supervisor: Prof. Pauli Murto

Advisor: Asst. Prof. Ciprian Domnisoru

This thesis aims to first provide basic understanding of directed acyclic graph causal inference and then apply this knowledge on how it is affecting its applicability to econometric research. First part is going through theoretical literature on graphs and statistical graphical models. This thesis will especially concentrate on model developed largely by Judea Pearl with its identification machinery. Then the theory is used to show some specific features considering the framework and taking a look at pre-existing literature in economics regarding this kind of model. Finally, some benefits and possibilities currently as well as in future and some possible developments and their effects for possibilities in economics are discussed.

This part concludes that there is quite a lot of challenges considering these models such as problems with instrument variables as well as with other applications requiring shape restrictions for functions used. However there also already exist use cases such as finding good control variables and techniques that allow more flexibility to used data, like the methods based on selection nodes.

Keywords: directed acyclic graphs, causal inference, nonparametric methods

Contents

Abstract	ii
Contents	iii
Symbols and abbreviations	iv
1 Introduction	1
2 Overview of directed acyclic graphs	3
2.1 Directed acyclic graph as structure	3
2.2 Graph operators	4
3 Directed acyclical graphs for statistics and causal inference	6
3.1 D-separation and independence	6
3.2 Statistical interpretation of directed acyclical graphs	8
3.3 Do-calculus and identification	9
4 Applications and properties of DAG framework in context of econometric analysis	14
4.1 Existing applications of DAGs in economic literature	14
4.2 Theoretical problems for economic applications	16
4.3 Advantages and possibilities of graphical models in econometrics	18
5 Conclusion	21
References	23
Articles	23
Books	26
Other sources	26

Symbols and abbreviations

Symbols

V	set of variables $\{V_1, V_2, \dots, V_N\}$
U	set of unobserved $\{U_1, U_2, \dots, U_N\}$
ϵ	set of edges $\{\epsilon_1, \epsilon_2, \dots, \epsilon_N\}$
\mathcal{G}	graph
\mathcal{Z}	subgraph

Operators

$\perp\!\!\!\perp$	independent
\nperp	not independent
$Pa(\cdot)$	parents of given node
$Ch(\cdot)$	children of given node
$De(\cdot)$	descendants of given node
$Pr(\cdot)$	probability

Abbreviations

CI	causal inference
DAG	directed acyclic graph
DCG	directed cyclic graph
SEM	structural equation model
LMC	local Markov condition
IV	instrumental variable
VAR	vector autoregression
ECM	error correction model

1 Introduction

Declaring causal inference (CI) has been for a long-time debated area of social sciences in general, and also in economics. It is naturally very interesting for scientists across the fields to establish more profound connections than correlation to events affecting their field studies. It also works as a way for economists to have larger effect on society as it enables policy advises to have more solid foundation on the effects of particular policy changes beforehand.

Purpose of this thesis is to give an overview of one of such methods that can be used to infer causal connections from data. Method discussed here, directed acyclic graph causal inference (DAG CI), is a method used in multiple fields like computer science and epidemiology, but have had smaller role in social sciences even though it has sparked some debate. However, in recent years there has been more interest also from these fields of study and one of the main advocates of this method, a computer scientist who has done research on this method, Judea Pearl is arguing that DAG methods should be used also in economics (Pearl, 2014), and he have been discussing with economists such as Guido Imbens over articles (Imbens, 2014; Imbens, 2020) and James Heckman and Rodrigo Pinto (Heckman and Pinto, 2015) about the suitability of the framework in economics.

First content section 2 tries to explain most essential set theory concepts that is needed for understanding graphs in this context and are prerequisites for understanding the statistical interpretation on DAGs. This part is mainly based on Pearls older paper going through the fundamentals of this method (Pearl, 1998) and textbook about CI (Peters, Janzing, and Schölkopf, 2017).

Second section 3 builds on top of the first and starts to give statistical features for graphs and going through some of the CI tools used with graphs. Properties of model and methods introduced here are needed later to understand why some specific parts of the model or tool associated with it are either useful or causing troubles in econometric applications. Two important topics here, developed largely by Pearl and his coauthors, are *do-calculus* and *identification methods* in more general sense. Literary considering these will contain multiple articles by Pearl for definitions of these tools. In addition, for that some contrasts are made with alternative choices that might be used like hypothetical graphs by Heckman (Heckman and Pinto, 2015) which introduces a bit differing way to use graphs in CI applications.

Last section (4) is solely dedicated for finding out the economics perspective, introductions of some techniques that can be used with model, and some more recent developments in these methods. Beginning of this section presents some of the existing empiric economic papers using at least some form of DAGs. These are macroeconomics articles considering Granger causality and thus not using

identification strategies offered by DAGs. This is also one of the problems of this thesis that there is not much empirical economic research that would yet be done with DAGs. Thus, it is mostly concentrated towards the possible benefits and shortcomings which are gone through more systematically in sections 4.2 and 4.3. These parts are mostly looking at some newer research and working papers considering some of the still existing problems for economics such as implementation of equilibrium condition and instrument variables as well as trying to come up with some of the current merits and features which are still under development as this is written but might affect the economic usability.

In the last part of the section 4, I try to find out how this model would benefit economic research especially in policy settings and features which might make it more accessible to conduct research in these cases compared to randomized controlled trials. Some of the discussed methods offer possibility to handle selection, transport model to different locations without necessarily estimating it again from the beginning and meta-analysis or synthesis approaches enabling combining information.

2 Overview of directed acyclic graphs

This section provides an overview of general structure of graphs as well as what kind of notations are used with graphs in general and with statistical interpretation. First the basics of graphs as structure and terms used as a basis for the statistical interpretations are gone through. This should provide necessary concepts needed to understand the framework as well as at later stage help understand the features either causing problems when DAGs are used in economic settings or features that might benefit economic analysis.

Secondly the operators used to define further properties that make graphs useful in statistical applications are introduced. These properties are very generally applicable for graphs also outside the CI framework and can be given for graphs as a general mathematical object. Implications of those are highly important for further discussion and to understand the building blocks of the statistical model as well as defining features such as independence in model.

2.1 Directed acyclic graph as structure

DAGs are at simplest just set containing variables V that have some extra restrictions regarding the way variables, or nodes in graph context, are connected withing the graph. Variables in V themselves do not have many restrictions. Graphs in this use should not be mixed with another meaning of word in mathematics. Graph showing for example picture of some function on x and y-axis is entirely different thing.

Aforementioned restrictions for DAGs are defining the set of possible relations for variables in the graph in hand as well as defining not permitted structures in the topological paths. Paths build up from the edges contained in ϵ which contains the relational information of graph. Edges in DAGs thus have property of direction and information where edge begins and where it ends. Edges and directions of a graph are not symmetric relations in case of DAGs since $\epsilon \subseteq \{V^2 \setminus (i, i)\}$ where $\epsilon_i = (i, j) \neq (j, i)$ which is necessary but not sufficient condition for acyclicity. When the graph is directed it leads to the topological ordering necessarily being proceeding at every point and never having an edge leading backwards in topological order as figure 1 shows (Peters, Janzing, and Schölkopf, 2017).

In this thesis specially directed and acyclic graphs are discussed and thus edges have the direction information or the information of in which direction each edge can be travelled. Acyclicity means that one node cannot exist twice in any possible path. In case node would exist twice in path it could appear infinitely many times in path due to the nature of cycles. This radically reduces the number of

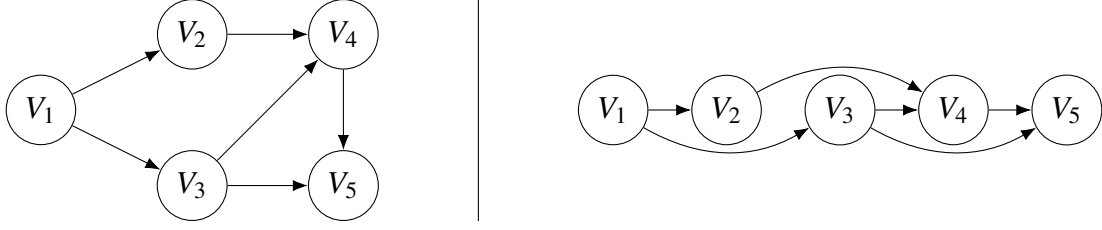


Figure 1: DAG and its topological ordering. If graph on left side would contain cycle it would be directed cyclic graph (DCG) and topological ordering would not be possible due to loop.

algorithms and methods that can be applied to the graph as well as the assumptions that can be made in the context of the graph allowing statistical interpretations. Breaking acyclicity also prevents use of the CI methods in similar form presented in this thesis. If there would exist cycles graph becomes directed cyclical graph and DCGs are fundamentally non recursive structures. This breaks some of the relationships that DAGs have, and which will be discussed in more detail and in economic context in section 4 (Heckman and Pinto, 2015). In other words, cyclicity introduces feedback mechanism for the structure which would lead to the graph being recursive and variables would therefore have causal effect with themselves. This is making the handling of model much harder, and methods introduced for directed acyclic graphs are not many times applicable in these cases at least without some modification or generalization of those. Theory for these different structures is not discussed in this thesis in depth, but some of the properties for those models are discussed in section 4 due to their high significance or possible significance for economic applications.

2.2 Graph operators

As graphs have additional conditions defined compared to normal set of objects there also exists operators that graphs have specially defined for them. These operators are also necessary for the causal analysis as these give tools for analyze and define parts of graph and differentiate subgraphs with specific features from the rest.

First concept to know is a path. Path in directed graph consists of variables that have consecutive edges connecting them so that for every variable in path, $V_i \rightarrow V_{i+1}$ holds. These also have role in CI as they describe the routes that causal relationships are influencing other variables. Also tightly connected to paths are the colliders, which always exists in relation to paths. Collider in path is structure where aforementioned condition does not hold but there exist directed edge so $V_i \rightarrow V_{i-1}$. Example of such can be seen in figure 1 DAG where path from V_1 to V_3 has subgraph $V_1 \rightarrow V_3$ where no collider exist.

Two of the other paths clearly have colliders as $V_1 \rightarrow V_2 \rightarrow V_4 \rightarrow V_5 \leftarrow V_3$ has V_5 forming collider and $V_1 \rightarrow V_2 \rightarrow V_4 \leftarrow V_3$ has V_4 as collider node.

Path such as $V_2 \leftarrow V_1 \rightarrow V_3$ in figure 1 are still further called fork. In fork middle node has two edges emerging from it. Path such as $V_1 \rightarrow V_3 \rightarrow V_5$ are called chains and in this type of path all edges are parallel.

With paths it is possible to define further relations, usually called kinship, for nodes or subgraphs. These operators are also needed to analyze causal structure in graphs since those enable distinguishing independence in graph and thus to define the statistical estimates for causality. Ancestor of variable V_i in directed graph, is such other variable appearing before V_i in topological order that has direct path to V_i . If node is appearing only one edge before V_i it is called parent node and set of parents is denoted as $Pa(\cdot)$. Respectively if there is direct path from V_i to some other node it is called its descendant and set of descendants is denoted by $De(\cdot)$. If there is path from V_i , only one edge long, node is called children $Ch(\cdot)$.

3 Directed acyclical graphs for statistics and causal inference

Graphs need or at least most of the time are used with, some special mathematical machinery when those are applied for CI. This section will take an overview to statistical applications of graphs and especially Markovian or causal graphs by going through some of the most fundamental tools to understand for CI in graphical models on conceptual level.

First part will be defining the way DAGs bring together necessary conditions for conditional independence and provides one of the most crucial concepts for the probabilistic interpretation, local Markov condition. With d-separation it is possible to isolate the necessary variables by conditioning the right set of variables and with the combination of these two we can start to make statistical sense within graphs.

With the help of local Markov condition, we can now factorize conditional probabilities out of the set of variables by conditioning the correct set of variables. For the first time everything talked about the general structure, operators and methods will come together and make it possible for the graph to make sense in statistical sense. At this point it becomes possible to define the model within which the causal analysis is going to operate.

Third part then introduces the ways to identify causal effects between variables and few of the most common strategies for that. In addition, do-calculus which is distinguishing DAGs from other causal frameworks and having its impact on the popularity of the model. This part will cover the mathematical machinery and the rules associated with it after which main components of the DAG framework in its simplest form is covered.

3.1 D-separation and independence

D-separation is only used in context of graphical models and implies independence between two variables within graph as it is the condition set by which independence can be decided. These conditions are shown in table 1 which also illustrates how d-separation ties the concept of independence into graphs. Variables can be independent by two ways in graph. First one is if only connecting paths are colliders. That leads two variables being independent without any further action or conditioning. Second way to establish d-separation is to have ordered pair of variables (X, Y) be separated from each other by set of variables \mathcal{Z} . Set of variables \mathcal{Z} however have few extra conditions for the variables it is containing:

1. Variables in \mathcal{Z} are in path between variables X and Y in paths that are either chains or forks.

This would correspond controlling $\{V_2, V_3\}$ or $\{V_2, V_4\}$ in figure 1 if the exposure variable is V_1 and outcome variable V_5 .

2. Variables of \mathcal{Z} are not variables or descendants of variables that construct collider on path.

From above can be seen there can exist multiple efficient¹ subsets of \mathcal{Z} even though also conditioning all other variables than colliders or their descendants would also lead to proper d-separation (Pearl, Glymour, and Jewell, 2016). These rules are also illustrated by table 1. However choosing good set of variables to control instead of all of \mathcal{Z} can also be utilized in situation where for example measuring some variable would be harder than others. Further, d-separation can help us make testable hypotheses that are implied by the graphical model. This is something setting graphical models apart from other types of CI methods by its very mechanical character and also key part in enabling causal claims from observational data.

Structure	Implied independence	
	Independence	Conditional independence
Directed path: $X \rightarrow \mathcal{Z} \rightarrow Y$	$X \not\perp\!\!\!\perp Y$	$X \perp\!\!\!\perp Y \mathcal{Z}$
Fork: $X \leftarrow \mathcal{Z} \rightarrow Y$	$X \not\perp\!\!\!\perp Y$	$X \perp\!\!\!\perp Y \mathcal{Z}$
Collider: $X \rightarrow \mathcal{Z} \leftarrow Y$	$X \perp\!\!\!\perp Y$	$X \not\perp\!\!\!\perp Y \mathcal{Z}$

Table 1: Independence in DAG. X denotes the exposure variable and Y output variable.

To have independence with these conditions, model should also be Markovian and fulfill local Markov condition (LMC). Model is considered Markovian when all the unobservable variables or residuals are jointly independent, and the model is acyclic. However sometimes some of those unobservables variables are showed in graph and if we know which variables it is affecting i.e., which variables are having some jointly dependent unobservable variables. By including unobservables making variables jointly dependent, it is still possible to work with it. For the causal claim to be testable some extra requirements need to be taken care of while choosing \mathcal{Z} as those unobservables open new paths that need to be conditioned or in case of collider not conditioned. These kinds of models are called semi-Markovian.

¹Set containing minimum amount of variables that fulfills the conditions for d-separation.

Local Markov condition on the other hand is the condition that is needed to provide the independency by d-separation and thus also translate the graphical properties of a graph to causal claim (Heckman and Pinto, 2015). LMC is telling us that the variables are independent of other ancestors they have, conditional on their parents. Equation 1 is saying that all v_i in V are independent of the set of variables containing its other ancestors than parent if conditioned with its parent. Formally defined LMC is:

$$\forall v_i \in V: v_i \perp\!\!\!\perp V \setminus \{De(v_i) \cup \{v_i\}\} \mid Pa(v_i) \quad (1)$$

Now with the LMC condition and graphoid statements that define the independence conditions, and were found out to hold in undirected graphs (Paz and Pearl, 1985) and later in directed (Pearl, 1986), the concept of independence starts to make sense in graphical context. These conditions need to be met for conditional independence to hold in graphs. With these it also gets possible to start to use DAGs for statistical purposes.

3.2 Statistical interpretation of directed acyclical graphs

From LMC (1) it is finally possible to derive the general factorization of variables in graphs. As the LMC suggests the parents of node are somewhat special group in context of the probabilistic interpretation by saying that other ancestors than parents are in fact irrelevant for the factorizing of probability distribution for variable. This is same as with any probability calculation but as mentioned in section 3.1 it is also found to be applicable for graphs. Joint distribution for variables $Pr(V_1, V_2, \dots, V_n)$ in recursive model can be derived from LMC (1). With $|V| = N$ variables that are ordered as V_1, V_2, \dots, V_{n-1} not being descendants of V_n and V_{n+1}, \dots, V_N being descendants of the same variable V_n , implying $Pa(V_n) \subseteq V_1, V_2, \dots, V_{n-1}$ (Heckman and Pinto, 2015). With this and assumption regarding exogenous variables U to be jointly independent, it is possible to show conditional independence being:

$$\begin{aligned} Pr(V_1, V_2, \dots, V_n) &= \prod_{V_n \in V} Pr(V_n | V_1, V_2, \dots, V_{n-1}) \\ &= \prod_{V_n \in V} Pr(V_n | Pa(V_n)) \end{aligned} \quad (2)$$

With factorization defined it gets possible to give some meaningful causal interpretation properties for graphs as probability is now meaningful concept in context of graph.

To start making sense of the model with data and having causal interpretation for the relations of functions between nodes those need to be defined. There have been and is multiple different approaches

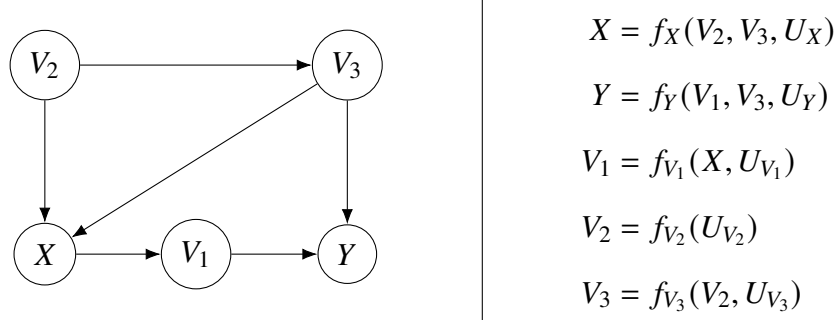


Figure 2: Non-parametric SEM with the graphical interpretation of the model.

for this. One of the first approaches was to write SEM with linear equations like $y = \beta x + u$ and then give the causal relation in graph based on the theoretical or observed knowledge of situation (Wright, 1921). This can however be generalized also to non-parametric functions which is for now on thought to be the default model here. SEM is constructed so that it contains functions for all observed variables (Pearl, 2008), however in model of figure 2 each function has all of the unobserved variables specific to variables explicitly written to make it a bit more clear even though those are not shown in the figure.

3.3 Do-calculus and identification

Pearl's approach includes the special notations and tools of do-calculus as he calls it. Do-calculus rules could be thought to be similar in principle to other algebraic expressions. By manipulating some expression with unwanted notations, in this case the do-notation, with the rules included in do-calculus we can get rid of those. This enables estimation with regular tools of probability theory as equations 9 & 10 will demonstrate. Thus, it is just set of axioms which are applicable only in context of graph and changes do-statements to regular probabilities. One of the nice properties of do-calculus is that, originally Pearl only conjectured that only the three rules would be sufficient to find causal connections. More recently it has been proven that, indeed these rules are a complete system for finding causal connections (Shpitser and Pearl, 2006; Huang and Valtorta, 2006). There are also rules that can be applied with less restrictive rules (Hyttinen, Eberhardt, and Järvisalo, 2015), but these approaches won't probably be as useful with economic applications as they are in machine learning applications. Idea with these rules is to apply those before declaring causal structure with even less strict assumptions about the underlying causal structure, while economists usually tend to build models based on some theoretical framework.

To get familiar with do-calculus and the identification methods developed for DAGs let's first

introduce few new notations. When stating independence as displayed in table 1 we might want to give some further conditions under which the independence is realized. If we have graph \mathcal{G} and within that variables X, Y, Z and W we can use notations such as $\mathcal{G}_{\overline{XZ}}$ or $\mathcal{G}_{\overline{XZ(W)}}$. Variable with overline in subscript of graph tells that all the incoming arrows for this variable are blocked. Respectively with underline it denotes all the appearing arrows from variable are blocked. When notation is used as done here with $\overline{XZ(W)}$ it denotes all Z -nodes, not ancestor of W -nodes in $\mathcal{G}_{\overline{X}}$ are blocking incoming edges. With these we can now read do-calculus rules.

1. Insertion/deletion rule for observation:

$$Pr(y|do(x), z, w) = Pr(y|do(x), w) \text{ if } (Y \perp\!\!\!\perp Z|X, W)_{\mathcal{G}_{\overline{X}}} \quad (3)$$

2. Action/observation exchange:

$$Pr(y|do(x), do(z), w) = Pr(y|do(x), z, w) \text{ if } (Y \perp\!\!\!\perp Z|X, W)_{\mathcal{G}_{\overline{XZ}}} \quad (4)$$

3. Insertion/deletion of action:

$$Pr(y|do(x), do(z), w) = Pr(y|do(x), z, w) \text{ if } (Y \perp\!\!\!\perp Z|X, W)_{\mathcal{G}_{\overline{XZ(W)}}} \quad (5)$$

List 1: Do-calculus rules (Pearl, 2009)

There exists at least two different formulation for these rules (Pearl, 1995; Pearl, 2009) of which I think these are the cleaner one and easier to understand as well as more used in the literature.

The reason why this set of rules exists is to uncover other causal connections not identifiable with these three simpler schemes (6, 7 & 1). Do-calculus enables us to make sense of more complicated situations and find all causal connections due to the completeness (Pearl, Glymour, and Jewell, 2016). It also packs these methods in algorithmic form, making identification easy for computer. Identification methods can still be applied to graph without any modification in situation where necessary conditions are met from the beginning. For example graph in 2 is already fulfilling conditions for *the backdoor criterion*.

Backdoor Criterion: In relation to ordered pair (X, Y) in \mathcal{G} , set of variables Z satisfies backdoor criterion if there is no descendants of X in Z and Z blocks all paths between X and Y containing arrow to X . When \mathcal{G} satisfies these all of above mentioned criteria, causal effect is attained by

backdoor adjustment:

(Pearl, Glymour, and Jewell, 2016)

$$Pr(Y = y|do(X = x)) = \sum_z Pr(Y = y|X = x, Z = z)Pr(Z = z) \quad (6)$$

This identification method can be applied to graph in figure 2 giving causal effect of X to Y . In this graph the adjustment set of variables, given by backdoor criterion, is only one variable $\{V_3\}$ and needs to be conditioned. By doing so and using the causal effect equation 6 it can be estimated as $\sum_{v_3} Pr(y|v_3, x)Pr(v_3)$ which is giving averaged joint distribution with V_3 conditioned. Other criteria that could be used are:

Frontdoor Criteria: Set Z satisfy frontdoor criterion relative to (X, Y) when:

1. Z intercepts all of the paths from X to Y .
2. No unblocked paths from X to Z .
3. All backdoor paths from X to Y are blocked by X .

(Pearl, Glymour, and Jewell, 2016) in this case causal effect can be identified by adjustment formula:

$$\begin{aligned} Pr(Y = y|do(X = x)) \\ = \sum_z \sum_{x'} Pr(Y = y|Z = z, X = x')Pr(X = x')Pr(Z = z|X = x) \end{aligned} \quad (7)$$

Z-identification: X, Y, Z being disjoint set of variables and \mathcal{G} the causal graph containing those. Causal effect $Q = Pr(y|do(x))$ is zID or z -identifiable in \mathcal{G} if one of next conditions holds: (Bareinboim and Pearl, 2012a)

1. Q is identifiable in \mathcal{G} (this must trivially hold if identification is done).
2. There exists $Z' \subseteq Z$ such that next holds:
 - i. X intercepts all directed paths from Z' to Y .
 - ii. Q is identifiable in $\mathcal{G}_{\overline{Z'}}$.

These criteria are in essence very similar to backdoor criterion on applicational level but do-calculus works bit differently as those rules can be applied multiple times to acquire the estimator. Graph in figure 3 for example has two backdoor paths going through V_1 as there exists now this unobservable U_1 . Causal effect could be calculated by conditioning on V_1 . Note that even though figure 3 is showing also

Heckman's hypothetical model all of the calculations are done with Pearl's do-calculus notation and right side of the model is not used in this example. In Heckman's framework this would happen with tools provided by standard probability theory.

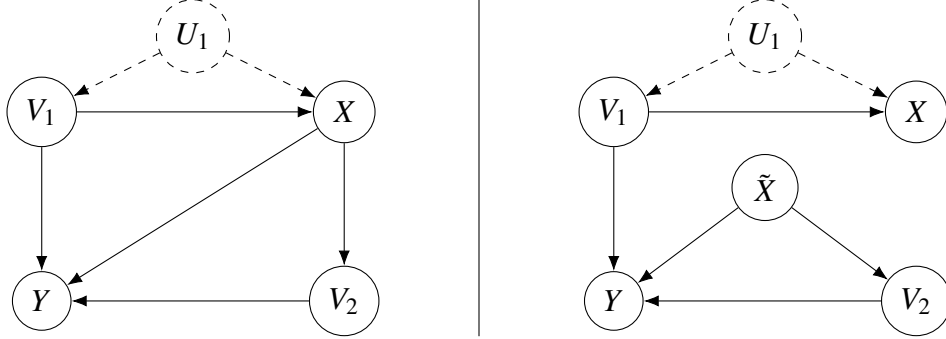


Figure 3: Graphs with two backdoor paths. Graph on left is showing empirical graph on which Pearl's model operates. Right is Heckman's addition to model showing hypothetical model where value of \tilde{X} is added and edges emerging from X are moved to emerge from that.

$$Pr(y|do(x)) = Pr(y|do(x), v_1)Pr(v_1|do(x)) \quad (8)$$

This however, by itself is not sufficient expression to be actually estimated. However it is still possible to manipulate expression 8 with do-calculus so that it can be estimated.

$$\sum_{v_1} Pr(y|do(x), v_1)Pr(v_1|do(x)) = \sum_{v_1} Pr(y|x, v_1)Pr(v_1|do(x)) \quad (9)$$

$$= \sum_{v_1} Pr(y|x, v_1)Pr(v_1) \quad (10)$$

First equation 9 is using do-rule 2 (4) and second equation 10 uses do-rule 3 (5). Even though at first glance it might seem the setup here is not identical to those as both rules contain both "W and X-variable" from do-calculus rules 1, which do not exist in this model. By testing the given conditions, it appears that both of those are actually met and those can be applied.

Rule 2 is first applied to expression 9 and it is applicable since placing variables from figure 3 to independence condition yields independence $(Y \perp\!\!\!\perp X|V_1)_{\mathcal{G}_{\tilde{X}}}$. By this, do-operator of expression 8 can be removed making it $Pr(y|x, v_1)$. Expression 10 is applying do-rule 3. Again by placing variables from this model the condition is holding as $(X \perp\!\!\!\perp V_1)_{\mathcal{G}_{\tilde{X}}}$ and enables the modification suggested by rule 3 to be applied for last part of expression 10 so it can be wrote as $Pr(v_1)$.

This expression would now enable evaluation of causal effect since it does not contain do-operators anymore. With do-calculus it would be even possible with observational data. CI with observational is somewhat controversial topic on which Pearl is strong advocate of it (Pearl, 2018). Despite that, Imbens writes economists are many times believing randomized controlled trials to be more convincing. Or alternatively if doing inference with observational data using research design modeled as such, like regression discontinuity design, and taking advantage of situations where randomization is true for at least subpopulation of sample (Imbens, 2020). Problem with these methods is that these does not have such clear graph representation as Imbens points out.

In addition to Pearl's model there exists also other approaches, most significantly in context of economics system developed by economist James Heckman and Rodrigo Pinto (Heckman and Pinto, 2015). Their model uses approach, derived from Haavelmo's work regarding CI (Haavelmo, 1943; Haavelmo, 1944) and applied to graphs. Heckman's model is using, in addition to the empiric model common to both model, hypothetical model shown in figure 3. Difference in the hypothetical version is the added variable \tilde{X} that does not have any incoming edges but all the arrows that were emerging from X are now emerging from \tilde{X} . This new variable contains the values for which the causal effect is assessed. This differentiates models as Pearl's version is manipulating empirical model with do-calculus (1) and Heckman is considering entirely different graph for estimation that has one more node. Benefit from this is that in hypothetical model no other tool, such as do-calculus is needed, and model can be estimated with basic probability theory.

Heckman's model is making some of the identification tasks easier of which the IV-identification is probably most important to economics. Identifying IV is not possible in basic form of Pearl's framework. Even though This approach has some desirable features there does not exist much literature around this version on model. As far as I am aware there neither exist any empirical research using this model. For that reason, focus will be in Pearl's approach.

With tools provided in this section we can now understand the basics of identifying causal effects in graphs as well as start to see some flaws and benefits for specific applications. Something more to mention, connected to these methods is that since graphical CI methods have largely been developed by computer scientist, there exists rather efficient algorithms to execute each method introduced above in practice making those quite nice and easy to use, of course still bounded with the normal restrictions for algorithmic efficiency related to graphs.

4 Applications and properties of DAG framework in context of econometric analysis

This part of thesis will first go through some research papers wrote in economics using DAGs. This literature is not very broad and those papers that are written using this framework are only using it to find out correlation structures of time series data with Granger causality and not using causal identification methods.

After this, some theoretical problems that are preventing the use for applications like applied micro econometric where those features that really make it useful in other fields like do-calculus could be used. Here restrictions of functional form and structure appear as big obstacles to work out to have possibilities for wider use in econometrics.

Lastly there is section going through benefits of DAGs for economic applications. This section will take a bit closer look at some of the topics already touched before like easy execution of identification. In additions some new features for framework, like transportability are introduced and some possibilities with these are discussed.

4.1 Existing applications of DAGs in economic literature

Existing empirical economic literature using DAGs at least in some form is mostly concentrated around time series data and modeling phenomenon like price drivers or contagion and transmission channels in international trade and financial applications. These papers are establishing Granger causality in chosen model (Awokuse and Bessler, 2003; Bessler and Yang, 2003; Yang, Guo, and Wang, 2006). There are also some more recently published articles covering topics such as natural gas price drivers (Ji, Zhang, and Geng, 2018), this also use similar model to those in older papers. The above-mentioned papers are using DAGs in very similar manner with similar techniques as introduced in sections 3.1 and 3.2. As these are establishing Granger causality with time series macro data, identification methods are not used due to these graphs not being causally sufficient.

From methodological viewpoint, and especially DAG viewpoint these articles are all using quite similar methods and I have picked one of those that represents well this literature to go through more thoroughly. Yang's article is using VAR error correction model (ECM) on which it is applying DAG. ECM is model used for cointegrated time series, and in this case unit root of CPI is tested by the Johansen's trace test (Johansen, 1991) and deemed significance. ECM enables more precise analysis of cointegration structure and dynamics between short- and long-term effects differentiating it from

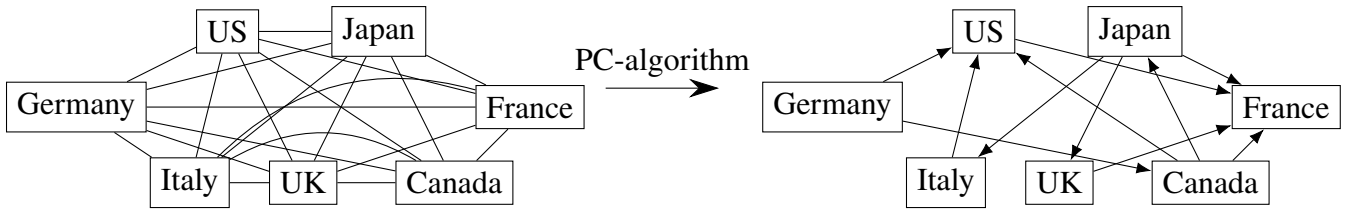


Figure 4: Recreation of graph from *International transmission of inflation among G-7 countries: A data-determined VAR analysis Jian* (Yang, Guo, and Wang, 2006). First graph is complete or has all edges on place while second shows the graph after PC-algorithm with $p = 5\%$.

normal VAR model (Lütkepohl and Krätzig, 2004). With results of this model Yang et al. calculates forecast error in this model and uses covariance decomposition of that as input for DAG. Other articles mentioned are having setups very close to or even identical to this and applying it to data considering their research question.

DAG part in each of the papers are executing PC-algorithm (Spirtes, Glymour, and Scheines, 2000) which is identifying the statistically significant edges in graph. In principle it works by testing whether, for each $V_i \in \mathbf{V}$ independence $V_i \perp\!\!\!\perp V_j$ where $V_j \in \{Pa(V_i) \cup Ch(V_i)\}$ when $i \neq j$ and $V_i, V_j \in \mathcal{G}$ with chosen statistical test holds, in this case Fischer's z-statistic. If $V_i \perp\!\!\!\perp V_j$ holds significantly then algorithm proceeds to removing the edge between them. Then it goes on with the same setup, but adding conditioned variables one by one as $V_i \perp\!\!\!\perp V_j | \{V_1, V_2, \dots, V_n\}$ when $V_1, V_2, \dots, V_n \in \{\mathbf{V} \setminus \{V_i\} \cup Pa(V_i) \cup Ch(V_i)\}$ and deleting edges if this holds statistically significantly. After the information of independence is stored, it is possible to deduce directions edges. In this case PC-algorithm can identify all edges with $p = 5\%$ but this is not necessarily the case and there might be multiple graphs in some Markov equivalent class. Markov equivalence class is set of graphs with same independence relations. In this particular case all of the directions are found out and Markovian class is singleton containing only graph shown in figure 4.

Articles discussed here are all operating with macro level data establishing Granger causality. Even though the name is suggesting causality, it is more of a hypothesis test on whether some time series is useful for forecasting other. Writers also acknowledge the possibility this is not complete graph of this process. Causal sufficiency which requires the graph to be complete description of variables affecting phenomena is an assumption that needs to be satisfied to interpret graph as causal. If graph is not sufficient for causal interpretation it is not possible to calculate any causal effects either. Thus, none of the articles is applying identification methods to calculate that. This is not really the type of use cases Pearl or Heckman and Imbes are talking in their text, as they are paying more attention to causal

identification.

There does not seem to be applied microeconomic papers using DAGs while those would be the type of research that could use identification methods. These studies, using only PC-algorithm are using DAGs for econometrics but not really doing the causal inference part all the way through. To do the causal inference part they would need to further try to find effect of some variable to other and estimate it, which in this setup would not be possible or at least credibility of those estimates would be questionable as these graphs are not fulfilling the causal completeness.

4.2 Theoretical problems for economic applications

There is not any one thing preventing the use of DAGs in econometric analysis. Neither do I believe that Imbens on his text about *The Book of Why* (Pearl and Mackenzie, 2018) or Heckman (Heckman and Pinto, 2015) are, despite quite critical view of models fit for econometric use, to say that there would not be any use cases for it in economic literature. However, DAGs still contain some inconveniences or assumption not that well applicable to economics which will be discussed next.

Some of the problems of DAGs are more technical in nature and others more of a question of taste and acquired habits, like which framework one thinks is clearer on its notation's, potential outcomes or DAGs. I will first look on the more technical side of these problems and discuss where DAG approach falls short of techniques used for econometrics. After that some of conceptual or otherwise inconvenient properties will be revised. It should also be noted that DAGs or at least the formal notations of do-calculus are relatively new techniques and the research regarding those is very active so some of the problems mentioned here might not remain unsolved for very long as the research proceeds.

One very clear lack in DAGs is that those cannot identify nonparametric IV-setup with methods introduced before. This is outcome of the fact that graphs are usually used as nonparametric models and the functional form of phenomena modeled is unknown from the beginning. The fact that the functions are not restricted can, and especially Pearl would in general argue to be good thing and add flexibility to the model. However, in IV setup this can cause some additional headache as it also leads to it not having as straight forward way to set shape restrictions for functions, like monotonicity in this case as Imbens points out (Imbens, 2020). There exists workarounds and other solutions like getting bounds for average causal effect which is possible in nonparametric setting (Balke and Pearl, 1997). There is also active research considering techniques for identifying instrumental variables in graphs as well as other identifying methods in nonparametric settings (Freyberger, 2017; Freyberger and Horowitz, 2015). One very recent working paper (Hoveid, 2021) which show a way to identify IV, but this is applicable

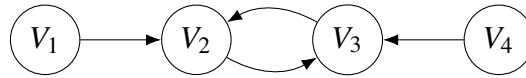


Figure 5: Graph with equilibrium structure making it DCG.

only under linearity. Also Pearl has noted about IV, that if variable in question can be reasoned to really be exogenous DAG might not be the right tool for the job (Pearl and Mackenzie, 2018). Lack of easy way for IV is said to be one of the big reason why DAGs are not adopted to economics by Imbens and Heckman (Imbens, 2014; Imbens, 2020; Heckman and Pinto, 2015) and they are spending quite large amount of the analyses talking about that and the function restrictions.

This problem also extends further than instrumental variables as in economics shape restricted functions are commonly used to bring some already known properties into model. Other occasions for restrictions might be even more bounding as it is common for those restrictions to be consequence of economic theory. This situation might arise for example while defining production function which may be desired to be monotonically increasing but with a diminishing return to scale such as Cobb-Douglas function. Economic theory is playing important role as substantive knowledge of underlying causal structure. After all causal claims need to be based on something more than just statistical relationship — making PC-algorithm imperfect way to gain knowledge about causality. Any correlation-based method cannot gain certain knowledge about causality, as correlation by itself does not imply causality and vice versa lack of correlation does not necessarily imply lack of causality.

Another phenomenon not easily represented in DAG is equilibrium condition, unfortunately something quite crucial in economics. This is immediate result of the definition as acyclic. Equilibrium is clearly resulting to cycle in graph as in figure 5, showing simple equilibrium condition. Equilibrium introduces feedback mechanism between, for example price and quantity constructing the cycle. This is also noted in other disciplines as well, and thus also this is something under very active research. There has already been some research on cyclic graphs that have been able to prove linear-SEM leading to global Markov property, which is generalization from local Markov property extended to cover, in addition to variables also subsets of variables in directed graph (Spirtes, 1994).

Some of the most recent, and quite ambitious try in a sense that it has tried to solve both problems of cyclical components in graphs and latent confounders, are made by Forré and Mooij (Forré and Mooij, 2017; Forré and Mooij, 2018; Forré and Mooij, 2019). This approach retains the Markovian properties for graph and enables non-linear functional forms. Forré and Mooij are proposing two different models that they are calling modular-SEM (mSEM) and more recent conditioned version of that input/output-SEM (ioSEM). These would take care at least some of those problems although with

the cost of making the model notably more complicated. Basic idea for both models is to generalize methods that were presented before by introducing these new class of graphs like hyperedged² directed graphs. Forré and Mooij are able to generalise methods like d-separation used in normal DAG to what they call σ -separation (Forré and Mooij, 2017) which is having same effect for these types of graphs but which can be considered as extension of d-separation as it reduces to effectively same method as d-separation when applied to DAG. By doing these changes Forré and Mooij are able to generalize the separation algorithm providing representation for conditional independencies in a model which includes cycles as well as conditions for applying mSEM with the graphs (Forré and Mooij, 2017; Forré and Mooij, 2018). And then make even further modification to rules so that any distributions and nonlinear functional forms are applicable as well as generalized versions of identifying algorithms (back door (6), front door (7), & zID -methods) as well as generalized version of do-calculus (1) (Forré and Mooij, 2019) although it should be noted that this work is partly still at the working paper stage rather than fully published. This kind of modified models, if those can be well implemented, could make a big difference for economic applicability as it would allow modeling equilibrium and enable using these in wider range of causal questions.

4.3 Advantages and possibilities of graphical models in econometrics

Despite the negative aspects, graphs hold potential for future and some already existing benefits for CI. Among those are easy visualization and estimation as well as some of the more technical properties like formalized way to combine or move models. This section will take deeper look at these.

Economic papers are usually not written just to be read by other economists and CI is many times done in settings which also have policy implications. Especially in those cases it might be beneficial to be able to present models and variables with their connections as simply as possible. For this DAGs are very clean and intuitive way. However, this comes with the cost of being very simplified version of what is happening so argument also goes other ways around and someone with lots of knowledge with statistical methods might find algebraic expressions more explicit and precise way to communicate models as Imbens notes (Imbens, 2020).

DAGs could also be used to find appropriate controls for the estimated model. Cunningham in his book (Cunningham, 2021) is making the remark that as table 1 shows when nodes are forming collider, conditioning it can open path causing dependency between variables. This would cause bias to model. To perceive the colliders DAGs are very good framework to think of as they show the concept of path

²Hyperedged graph or hypergraph is more general type of graph allowing edges to connect multiple nodes.

in much more evident way compared to regression function. Example of this can be seen in figure 1 where colliders are easy to see from graph (V_4 and V_5) but even easier from the topological order where all nodes with more than one incoming edge are colliders. This is also one of the areas which is under active research and for example this very recent working paper by Hünermund and his coauthors (Hünermund, Louw, and Caspi, 2021) are considering choosing controls for model with micro level data. This, still very early-stage paper, is developing system by which these bad controls could be systematically assessed. In addition to developing system and using it with simulated data they are using this for empiric data from paper by Blau and Kahn (Blau and Kahn, 2017) about gender wage gap. Hünermund et al. are testing the effect of controlling marital status to model, which is thought to be bad control due to endogeneity with women's decision to join labor-force. They are getting fairly similar results between their model when their model is including broader set of controls and OLS but finding out that the inclusion of marital status to widen the pay gap between models by 10.6% and concluding that choosing controls can significantly change the implications of study.

When estimating model with DAGs this clarity also extends to the estimation. Proving very complicated structural model with many variables can quickly get very complex. In DAGs proving is quite easy to show with identification methods. Many of the identification methods also have complete algorithms that can do the job, like the algorithm to find efficient sets for back door criteria (6) (Correa and Bareinboim, 2017). This saves the user of model from doing the more technical work and instead lets researcher to focus improving the causal structure and interpreting results. In general estimating models in this framework is quite easy as most of the new features are also developed with algorithmic solution from the beginning.

Another nice property of causal connections that are found in DAGs arise from the feature introduced as problem before but can also be beneficial in some cases. Very little restrictions for functional form in DAG, the exact problem causing pain with IVs lead also to those causal observations having very little restrictions in functional sense and thus holding with very little restrictive assumptions making it quite robust in that sense as Heckman points out (Heckman and Pinto, 2015). However, for economics this property is somewhat of a two-edged sword as it makes introducing other restriction in DAG model that might arise from economic theory.

One repeatedly occurring problem in economics is selection bias and for that there already exists machinery in DAG-literature that works without restricting the functional form that can be applied in model (Bareinboim, Tian, and Pearl, 2014). Selection node can be introduced to normal DAG and with help of that it is possible to expand the use of model for biased samples and use it with hybrid

sources where part of the estimates might be done with unbiased data and other with biased. This kind of situation might arise when some of the estimators that are necessary for the identification might be more easily available e.g., from some statistical authority, but other part would require using more unbalanced data. This could have potentially many applications and might reduce costs of doing research in this kind of situations. Example of such situation might be to help political decision making when question can be modeled as causal question, but relations have been very hard to establish or doing that would have been too expensive.

Selection nodes can also help make studies more externally valid or "transportable" as it is called in computer science literature (Pearl and Bareinboim, 2011). Term transportability is not exact synonym to external validity but applying same model to new place by changing some part of that with the technique established by Pearl (Pearl and Bareinboim, 2011). Conditions under which transportation can be applied are formally defined (Bareinboim and Pearl, 2012b). Term external validity is already used in some instances and in my opinion, even though not perfect, quite good to capture essence of the concept. Transportability is in a sense extension of selection nodes and in this case selection nodes are introduced to other parts of model where those are adding information about other domains of model that are some ways altered when model is used in other than its original location.

Third concept related to selection and transportability is meta-synthesis which enables integrating information from multiple models to one and thus estimating effects, with background information of the specific setup but with models originally empirically estimated somewhere else (Pearl, 2012). Meta-synthesis would enable a bit similar situation as transportability but rather than moving whole model, building a new one from pieces of multiple models. This could also benefit economic research on similar kind of situations as transportability as it might enable easier data gathering processes to areas where data is sparse or somehow hard to acquire.

These features could be divided to two different categories that might be described as the benefits of the conceptual framework and benefits that arise from the technical possibilities, concentrated around the selection nodes. In short the first category would contain the nice and understandable representation of model and the way through which other variables can be seen within that enabling bit different perspective for control variables. Second category adds selection bias recovery, transportability, and meta-synthesis to the toolkit of economist. These can help estimating model in places where data is sparse, it is other ways hard or impossible to get or just lower the cost of it when it might be too costly to estimate otherwise.

5 Conclusion

DAGs are already seeing lot of applications in areas such as epidemiology but have not been as popular in economics. So, what is the reasons for economists not to choose this approach for CI? I think this can be divided to roughly four categories that vary in difficulty of finding solution as well as on the nature of problem. First one is simply because it is quite new and have not been used in many studies. In other words, there is not that much economic research done with it which raises the bar for using it in new projects.

Second category contains technical questions, like functions shape restrictions. This is of course also partly guilty for the first one as it is one of the reasons that prevents use in some cases. However, for these problems there might be some solutions on a way or at least there are many people trying to come up with those as was discussed in section 4.2. Another technical problem is considering the trouble DAGs are having with the cyclicity as the name already suggests. This is making it more irrelevant and preventing the use of DAGs in econometric applications as many of those are considering some situation with equilibrium condition. Even though Pearl have tried to show this could be solved with fairly simple solution (Pearl and Mackenzie, 2018), for Imbens and Heckman this has not been (Imbens, 2020; Heckman and Pinto, 2015) sufficient solution since Pearls proposal is more or less just ignoring the equilibrium condition by making demand affect supply directly instead of through price.

Last of these general categories I identified relates to the way graphs are done and thought. As this approach is arising from computer science departments more than from economics departments in its modern form, it also contains some of the drawbacks and benefits discipline can provide. One of such, that is in principle a good thing is the ways in which identification is done. It is possible to save the essence of these methods to simple form and develop algorithm to do the job. Although there are also these machine learning methods in development, that are trying to do the identification automatically there must be some substantive knowledge about the situation. After that is done the rest is very mechanical and might feel very alluring to take answers as given even though this is completely conditional to the building the causal structure. This of course is not necessarily bad thing but rather something that needs lot of attention and questioning the assumptions model is making even though it might superficially look alright and it kind of lacks the derivation of model as these identifications are automatic. There also exists kind of mental asymmetry between adding and *not adding* edge in graph. Even though adding edges is of course important in sense that it builds the model by adding causal relationship, not adding edge is in a way even stronger assumption of the underlying phenomena. By not adding edge person doing the model is saying that variables are absolutely not affecting each

other, at least directly, in comparison to adding one might range from effecting just a bit to being the sole deterministic cause. In economics it is very hard to be convinced two events are not in any way connected. One of major mechanisms causing this being such a fundamental part of economics, namely the tendency for economic agents to somehow optimize their behavior in a way that is considering the surrounding environment and potential outcomes in given situation. This is making causal structures immensely complicated as the reaction to phenomenon's are dependent of such vast number of factors and full confidence for any given structure is hard to achieve. Not including edge in model could be compared to claiming that the instrument is exogenous in IV model. In DAGs it is not as strict as in some places it might not affect the model but regardless of that model could have bias if only one of the edges is missing so in principle there is $|V|^2$ minus added edges of those conditions to argue for. Some of those are also making model cyclical so these would prevent estimating causal effect.

Some of the interesting future research, apart from the theoretical challenges that needs to be solved for DAGs to find more use in economics could be related to the already mentioned (4.3) properties of selection, transportability, and meta-synthesis. Both theoretical question for generalizing results of model by identifying the differing factors in research setting, that have already gained interest of researchers (Cinelli and Pearl, 2021) and the empirical questions that might get some answers by applying those (Dahabreh et al., 2020) for transporting randomized controlled trial research to conclude average treatment effects in new target population. These also seems something that could offer lot for economic research and make it more accessible in some situations.

Such questions where these could add value for economic analysis, and through it to public decision making might be here in Finland questions regarding incoming reforms on social security and health care. Both reforms are allocating some of the responsibilities to provincial and municipal levels. However, it is not attainable to have randomized controlled trials considering all these units of governance due to costs associated. Still, it does not seem likely that those would all have homogeneous effects on these reforms. By doing research on more centralized manner and having broad data considering the selection within areas would allow using these models and some local, maybe even already existing registry data to get at least some approximates for the local effects on various questions. This could further extend ability to make informed decisions on lower levels of administration. In application these might enable answering questions like what kind of employment services would benefit each area most given the structure of population and other variables deemed relevant.

References

Articles

- Awokuse, Titus O. and David A. Bessler (2003). “Vector Autoregressions, Policy Analysis, and Directed Acyclic Graphs: An Application to the U.S. Economy”. In: *Journal of Applied Economics* 6.1, pp. 1–24. ISSN: 1514-0326. DOI: [10.1080/15140326.2003.12040583](https://doi.org/10.1080/15140326.2003.12040583).
- Balke, Alexander and Judea Pearl (1997). “Bounds on treatment effects from studies with imperfect compliance”. In: *Journal of the American Statistical Association* 92.439, pp. 1171–1176. ISSN: 1537274X. DOI: [10.1080/01621459.1997.10474074](https://doi.org/10.1080/01621459.1997.10474074).
- Bareinboim, Elias and Judea Pearl (2012a). “Causal inference by surrogate experiments: Z-identifiability”. In: *Uncertainty in Artificial Intelligence - Proceedings of the 28th Conference, UAI 2012*, pp. 113–120. arXiv: [1210.4842](https://arxiv.org/abs/1210.4842).
- (2012b). “Transportability of causal effects: Completeness results”. In: *Proceedings of the National Conference on Artificial Intelligence* 1.August, pp. 698–704.
- Bareinboim, Elias, Jin Tian, and Judea Pearl (2014). “Recovering from selection bias in causal and statistical inference”. In: *Proceedings of the National Conference on Artificial Intelligence* 4.July, pp. 2410–2416.
- Bessler, David A. and Jian Yang (2003). “The structure of interdependence in international stock markets”. In: *Journal of International Money and Finance* 22.2, pp. 261–287. ISSN: 02615606. DOI: [10.1016/S0261-5606\(02\)00076-1](https://doi.org/10.1016/S0261-5606(02)00076-1).
- Blau, Francine Dee and Lawrence Kahn (2017). “The Gender Wage Gap: Extent, Trends, and Explanations”. In: *Journal of Economic Literature* 55.3, pp. 789–865. DOI: [10.2307/j.ctt1tm7gsm.15](https://doi.org/10.2307/j.ctt1tm7gsm.15).
- Cinelli, Carlos and Judea Pearl (2021). “Generalizing experimental results by leveraging knowledge of mechanisms”. In: *European Journal of Epidemiology* 36.2, pp. 149–164. ISSN: 15737284. DOI: [10.1007/s10654-020-00687-4](https://doi.org/10.1007/s10654-020-00687-4).
- Correa, Juan D. and Elias Bareinboim (2017). “Causal effect identification by adjustment under confounding and selection biases”. In: *31st AAAI Conference on Artificial Intelligence, AAAI 2017* January, pp. 3740–3746.
- Dahabreh, Issa J. et al. (May 2020). “Toward Causally Interpretable Meta-analysis: Transporting Inferences from Multiple Randomized Trials to a New Target Population”. English (US). In: *Epidemiology* 31.3, pp. 334–344. ISSN: 1044-3983. DOI: [10.1097/EDE.0000000000001177](https://doi.org/10.1097/EDE.0000000000001177).

- Forré, Patrick and Joris M. Mooij (2017). “Markov Properties for Graphical Models with Cycles and Latent Variables”. In: arXiv: [1710.08775](https://arxiv.org/abs/1710.08775). URL: <http://arxiv.org/abs/1710.08775>.
- (2018). “Constraint-based causal discovery for non-linear structural causal models with cycles and latent confounders”. In: *34th Conference on Uncertainty in Artificial Intelligence 2018, UAI 2018* 1, pp. 269–278. arXiv: [1807.03024](https://arxiv.org/abs/1807.03024).
- (2019). “Causal calculus in the presence of cycles, latent confounders and selection bias”. In: *35th Conference on Uncertainty in Artificial Intelligence, UAI 2019*. arXiv: [1901.00433](https://arxiv.org/abs/1901.00433).
- Freyberger, Joachim (2017). “On Completeness and Consistency in Nonparametric Instrumental Variable Models”. In: *Econometrica* 85.5, pp. 1629–1644. ISSN: 0012-9682. DOI: [10.3982/ecta13304](https://doi.org/10.3982/ecta13304).
- Freyberger, Joachim and Joel L. Horowitz (2015). “Identification and shape restrictions in nonparametric instrumental variables estimation”. In: *Journal of Econometrics* 189.1, pp. 41–53. ISSN: 18726895. DOI: [10.1016/j.jeconom.2015.06.020](https://doi.org/10.1016/j.jeconom.2015.06.020). URL: <http://dx.doi.org/10.1016/j.jeconom.2015.06.020>.
- Haavelmo, Trygve (1943). “The Statistical Implications of a System of Simultaneous Equations”. In: *Econometrica* 11.1, p. 1. ISSN: 00129682. DOI: [10.2307/1905714](https://doi.org/10.2307/1905714).
- (1944). “The Probability Approach in Econometrics”. In: *Econometrica* 12.Jul, pp. iii–vi+1–115. DOI: [10.2307/1906935](https://doi.org/10.2307/1906935).
- Heckman, James J. and Rodrigo Pinto (2015). “Causal Analysis After Haavelmo”. In: *Econometric Theory* 31.1, pp. 115–151. ISSN: 14694360. DOI: [10.1017/S026646661400022X](https://doi.org/10.1017/S026646661400022X).
- Hoveid, Øyvind (2021). “Constructing valid instrumental variables in generalized linear causal models from directed acyclic graphs”. In: arXiv: [2102.08056](https://arxiv.org/abs/2102.08056). URL: <http://arxiv.org/abs/2102.08056>.
- Huang, Yimin and Marco Valtorta (2006). “Pearl’s calculus of intervention is complete”. In: *Proceedings of the 22nd Conference on Uncertainty in Artificial Intelligence, UAI 2006* July 2006, pp. 217–224. arXiv: [1206.6831](https://arxiv.org/abs/1206.6831).
- Hünermund, Paul, Beyers Louw, and Itamar Caspi (2021). “Double Machine Learning and Bad Controls – A Cautionary Tale”. In: 0, pp. 1–12. arXiv: [2108.11294](https://arxiv.org/abs/2108.11294). URL: <http://arxiv.org/abs/2108.11294>.
- Hyttinen, Antti, Frederick Eberhardt, and Matti Järvisalo (2015). “Do-calculus when the true graph is unknown”. In: *Uncertainty in Artificial Intelligence - Proceedings of the 31st Conference, UAI 2015*, pp. 395–404.

- Imbens, Guido W. (2014). “Rejoinder”. In: *Statistical Science* 29.3, pp. 375–379. ISSN: 08834237. DOI: [10.1214/14-ST5496](https://doi.org/10.1214/14-ST5496).
- (Dec. 2020). “Potential outcome and directed acyclic graph approaches to causality: Relevance for empirical practice in economics”. In: *Journal of Economic Literature* 58.4, pp. 1129–1179. ISSN: 00220515. DOI: [10.1257/JEL.20191597](https://doi.org/10.1257/JEL.20191597). arXiv: [1907.07271](https://arxiv.org/abs/1907.07271).
- Ji, Qiang, Hai Ying Zhang, and Jiang Bo Geng (2018). “What drives natural gas prices in the United States? – A directed acyclic graph approach”. In: *Energy Economics* 69, pp. 79–88. ISSN: 01409883. DOI: [10.1016/j.eneco.2017.11.002](https://doi.org/10.1016/j.eneco.2017.11.002). URL: <https://doi.org/10.1016/j.eneco.2017.11.002>.
- Johansen, Søren (1991). “Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models”. In: *Econometrica* 59.6, pp. 1551–1580. DOI: <https://doi.org/10.2307/2938278>.
- Paz, Azaria and Judea Pearl (1985). “GRAPHOIDS: A Graph-based Logic for Reasoning About Relevance Relations or When would x tell you more about y if you already know z”. In: p. 28. URL: https://ftp.cs.ucla.edu/pub/stat%7B%5C_%7Dser/r53-L.pdf.
- Pearl, Judea (1986). “Fusion, propagation, and structuring in belief networks”. In: *Artificial Intelligence* 29.3, pp. 241–288. ISSN: 00043702. DOI: [10.1016/0004-3702\(86\)90072-X](https://doi.org/10.1016/0004-3702(86)90072-X).
- (1995). “Causal Diagrams for Empirical Research Author”. In: 82.4, pp. 669–688. DOI: <https://doi.org/10.2307/2337329>.
- (1998). “Graphs , causality, and structural equation models”. In: *Sociological Methods & Research*. 27.2, pp. 226–284. DOI: [10.1177/0049124198027002004](https://doi.org/10.1177/0049124198027002004).
- (2008). “Causal Inference”. In: *Proceedings of Machine Learning Research* 6, pp. 39–58. ISSN: 20958099. DOI: [10.1016/j.eng.2019.08.016](https://doi.org/10.1016/j.eng.2019.08.016). URL: <http://proceedings.mlr.press/v6/pearl10a.html>.
- (2012). “The Do -Calculus Revisited.” In: *Proceedings of the Twenty-Eight Conference on Uncertainty in Artificial Intelligence* August, pp. 4–11. ISSN: 0000-0000. arXiv: [1210.4852](https://arxiv.org/abs/1210.4852).
- (2018). “Challenging the hegemony of randomized Controlled Trials: A commentary on Deaton and Cartwright”. In: April. DOI: [10.1016/j.socscimed.2018.04.024](https://doi.org/10.1016/j.socscimed.2018.04.024).
- Pearl, Judea and Elias Bareinboim (2011). “Transportability of causal and statistical relations: A formal approach”. In: *Proceedings - IEEE International Conference on Data Mining, ICDM* August, pp. 540–547. ISSN: 15504786. DOI: [10.1109/ICDMW.2011.169](https://doi.org/10.1109/ICDMW.2011.169).

- Shpitser, Ilya and Judea Pearl (2006). “Identification of joint interventional distributions in recursive semi-Markovian causal models”. In: *Proceedings of the National Conference on Artificial Intelligence* 2.July, pp. 1219–1226.
- Spirtes, Peter L. (1994). “Conditional Independence in Directed Cyclic Graphical Models for Feedback”. In: *Report CMU-PHIL-53*.
- Wright, Sewall (Jan. 1921). “Correlation and causation”. In: *Journal of Agricultural Research* XX.7, pp. 557–585.
- Yang, Jian, Hui Guo, and Zijun Wang (2006). “International transmission of inflation among G-7 countries: A data-determined VAR analysis”. In: *Journal of Banking and Finance* 30.10, pp. 2681–2700. ISSN: 03784266. DOI: [10.1016/j.jbankfin.2005.10.005](https://doi.org/10.1016/j.jbankfin.2005.10.005).

Books

- Cunningham, Scott (2021). *Causal Inference: The Mixtape*. Yale University Press. ISBN: 9780300251685. URL: <http://www.jstor.org/stable/j.ctv1c29t27>.
- Lütkepohl, Helmut and Markus Krätzig (2004). *Applied Time Series Econometrics*. Themes in Modern Econometrics. Cambridge University Press. DOI: [10.1017/CB09780511606885](https://doi.org/10.1017/CB09780511606885).
- Pearl, Judea (2009). *Causality: Models, Reasoning and Inference*. 2nd. USA: Cambridge University Press. ISBN: 052189560X.
- Pearl, Judea, Madelyn Glymour, and Nicholas P. Jewell (2016). *Causal inference in statistics : a primer*. West Sussex, England: Wiley, p. 159. ISBN: 1-119-18686-2.
- Pearl, Judea and Dana Mackenzie (2018). *The Book of Why. The New Science of Cause and Effect*. New York: Basic Books. ISBN: 978-0-465-09760-9.
- Peters, Jonas, Dominik Janzing, and Bernhard Schölkopf (2017). *Elements of Causal Inference: Foundations and Learning Algorithms*. The MIT Press. ISBN: 0262037319.
- Spirtes, Peter L., Clark Glymour, and Richard Scheines (2000). *Causation, Prediction, and Search*. 2nd edn, pp. 1–517. DOI: [10.1002/sim.1415](https://doi.org/10.1002/sim.1415).

Other sources

- Pearl, Judea (Oct. 2014). *Are economists smarter than epidemiologists? (Comments on Imbens’s recent paper)*. URL: <http://causality.cs.ucla.edu/blog/index.php/2014/10/27/are-economists-smarter-than-epidemiologists-comments-on-imbenss-recent-paper/>.